BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains

Michael L. Blinov, James R. Faeder, Byron Goldstein

and William S. Hlavacek

Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

June 8, 2004

ABSTRACT

Summary: BioNetGen allows a user to create a computational model that characterizes the dynamics of a signal transduction system and that accounts comprehensively and precisely for specified enzymatic activities, potential post-translational modifications, and interactions of the domains of signaling molecules. The output defines and parameterizes the network of molecular species that can arise during signaling and provides functions that relate model variables to experimental readouts of interest. Models that can be generated are relevant for rational drug discovery, analysis of proteomic data, and mechanistic studies of signal transduction.

Availability: http://cellsignaling.lanl.gov/bionetgen

Contact: bionetgen@lanl.gov

COMBINATORIAL COMPLEXITY

A problem that one confronts when attempting to model signal transduction is combinatorial complexity, which is caused by the many ways that signaling molecules can combine and be modified (Hlavacek et al., 2003). For example, a protein that contains n sites at which phosphate can be added or removed through the activities of kinases and phosphatases can occupy 2^n different phosphoforms. Adding further to this problem, post-translational modifications typically regulate the reversible assembly of heterogeneous signaling complexes, e.g., via protein-protein interactions that depend on phosphorylation. Even when only a few proteins are considered, as in a model for activation of the protein tyrosine kinase Syk (Faeder et al., 2003), the enzymatic activities, potential modifications, and interactions of the molecules imply a large number of possible molecular species, hundreds to thousands for systems we have considered. This complexity is unavoidable if we wish to develop predictive models that incorporate details at the level of molecular domains, the fundamental components of signal transduction systems (Goldstein et al., 2004).

RULE-BASED DOMAIN-ORIENTED MODELING

As part of our effort to study signaling by FcεRI, the high-affinity receptor for IgE antibody, we have developed a rule-based domain-oriented approach to modeling that addresses the problem of combinatorial complexity (Goldstein et al., 2002; 2004; Faeder et al., 2003; Hlavacek et al., 2003). In this approach, the possible states of molecular domains and rules for the activities and interactions of domains are specified. The rules

are then used in a computer program to generate a reaction network composed of all chemically distinct species and reactions implied by the specified properties of the molecular domains. An individual reaction is parameterized by the rate constant assigned to its class of reaction, each of which is defined by a rule. This approach to modeling is facilitated by BioNetGen, which allows a user to create multi-domain objects and specify reaction rules based on these objects through a text-based interface. Models, appropriate for chemical reaction kinetics in spatially homogenous reaction compartments, can be generated for a variety of systems.

SYNTAX OF MODEL SPECIFICATION

A BioNetGen input file defines 1) rate constants and concentrations; 2) molecular components, such as protein interaction domains and the potential states of these domains; 3) reaction rules, one for each type of reaction to be considered; and 4) output functions. The conventions of model specification are illustrated in Fig. 1. Sample input files are available at our web site, as well as a user's guide, a quick reference guide, and an online tutorial.

The molecular species in a model are specified as follows. A user can declare individual molecular species (Fig. 1(a)), multi-state species (Fig. 1(b)), and complexes composed of two multi-state species (Fig. 1(c)). An individual molecular species is declared by assigning it a name. A multi-state species declaration can be used to represent a protein that has a number of phosphorylation states or a scaffold protein that has a number of bound states as a result of interactions with multiple binding partners. A multi-state

species is declared by assigning it a name and specifying the number of possible states for each of the molecular domains to be considered. An individual species implied by the declaration of a multi-state species or complex is referenced by specifying its particular domain states. A set of species can be referenced by specifying a wild card (*) for the state of a domain. The components of a model must be declared as described above before they can be used in definitions of reaction rules and output functions. This requirement is imposed to prevent reaction rules from generating molecular species that are unanticipated by the user.

Reaction rules are written in the same form as a chemical reaction but apply to a range of reactants and products if they involve multi-state species or complexes and specifications of wild cards for domain states (Fig. 1(d)). A reaction rule generates a separate reaction for each set of reactants and products implied by its specification. These reactions are parameterized by the same rate constant(s). The validity of assigning the same rate constant(s) to a set of reactions is the responsibility of the modeler, who has the ability to specify particular domain states in reaction rules to account for steric clashes, cooperativity, and other factors related to the states of reactants that might influence the rate of a reaction. Thus, the user can define which components and modifications of a molecule or molecular assembly affect a particular chemical transformation and which do not. If a user assumes that only one or two domain states affect a given reaction, then the number of reaction rules (and rate constants) that a user must provide to specify a model is comparable to the number of molecular domains considered in the model, which is likely to be much less than the total number of reactions. The advantages and

disadvantages of this modeling approach have been discussed elsewhere (Hlavacek et al., 2003; Goldstein et al., 2004).

A user can define cumulative quantities that relate model variables to experimental readouts (Fig. 1(e)), such as the phosphorylation level of a particular protein. The ability to define such output functions is important because observable quantities typically reflect an ensemble of difficult-to-distinguish molecular species.

CAPABILITIES AND LIMITATIONS

BioNetGen, which is implemented in Perl, translates the high-level specification of a model, described above, into a chemical reaction network, i.e., a comprehensive list of the species and reactions implied by the user's declarations. The output can be read by other programs in the BioNetGen distribution, including a C program called Network that translates the list of reactions into a set of coupled ordinary differential equations (ODEs) and solves the ODEs using routines from the CVODE library (Cohen and Hindmarsh, 1996). Network sends the time-courses of concentrations and output functions in tabular format to files that can be imported into visualization software, such as Grace (http://plasma-gate.weizmann.ac.il/Grace), for which an interface is provided. BioNetGen also exports models in systems biology markup language (SBML) format (Hucka et al., 2003). As a result, models are usable not only by programs in the BioNetGen distribution but also by the various software tools that support SBML (http://sbml.org). These tools include not only ODE solvers but also programs that

implement discrete-event Monte Carlo algorithms for simulating stochastic chemical reaction kinetics (Gillespie, 1976).

The conventions of BioNetGen provide a concise language for specifying models that account for the modifications and interactions of molecular domains. For example, the input file that specifies the model of Faeder et al. (2003) consists of 95 declarations of parameter values, reaction rules, and output functions and requires seven kilobytes of memory. In contrast, the SBML file that specifies this model requires the explicit declaration of 3,680 unidirectional reactions and is more than a megabyte in size (because of both the verbose XML encoding and the number of reactions). BioNetGen may serve as a guide for the development of standards for representing and exchanging rule-based models in systems biology, which are currently being discussed and developed (Finney and Hucka, 2003; Franza, 2004).

We have used BioNetGen to generate models for early membrane-proximal signaling events triggered by antigen (Goldstein et al., 2002; Faeder et al., 2003), epidermal growth factor, erythropoietin, and interleukin-1 in mammals. We have also generated models for mitogen-activated protein kinase cascades involved in responses of yeast to α-factor pheromone and osmotic stress. These models are available at our web site, and they illustrate a range of BioNetGen capabilities.

Most software tools for modeling signal transduction require a user to make a declaration of some type for each species and reaction in a model, which is a severe limitation for

systems marked by combinatorial complexity. In contrast, BioNetGen interprets a small number of user-specified rules to generate a large reaction network. Rule-based generation of reaction networks is also facilitated by Cellerator (Shapiro et al., 2003), STOCHSIM (Le Novère and Shimizu, 2001), and other tools in development (see links to related software projects on our web site). An advantage of BioNetGen over the tools reported in the literature is the ability to handle aggregation of multi-state species, a critical feature of many systems (Goldstein et al., 2004). However, a general treatment of multi-component complexes will require further software development, because BioNetGen is currently limited to complexes of two multi-state species. Allowing three or more such species to aggregate requires additional inputs that significantly complicate model specification. Another limitation of BioNetGen at present is that it enumerates all possible species and reactions prior to simulation of the network dynamics. When the number of species is sufficiently large, it may be more practical to generate new species and reactions on-the-fly during a simulation (Hlavacek et al., 2003), which will require an integration of the rule evaluation and simulation capabilities. Extensions of BioNetGen are planned and will be announced on our web site.

ACKNOWLEDGEMENTS

This work was supported by grants GM35556 and RR18754 from the National Institutes of Health and by the Department of Energy through contract W-7405-ENG-36. We thank Ed Stites, Aileen Vandenberg and Jin Yang for beta testing.

REFERENCES

Cohen,S.D. and Hindmarsh,A.C. (1996) CVODE, a stiff/nonstiff ODE solver in C. *Comput. Phys.* **10**, 138-143.

Faeder,J.R., Hlavacek,W.S., Reischl,I., Blinov,M.L., Metzger,H., Redondo,A., Wofsy,C. and Goldstein,B. (2003) Investigation of early events in FcεRI-mediated signaling using a detailed mathematical model. *J. Immunol.* **170**, 3769-3781.

Finney,A. and Hucka,M. (2003) Systems biology markup language: level 2 and beyond. *Biochem. Soc. Trans.* **31**, 1472-1473.

Franza,B.R. (2004) From play to laws: language in biology. *Sci. STKE* **2004**, pe9.

Gillespie,D.T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.* **22**, 403-434.

Goldstein,B., Faeder,J.R., Hlavacek,W.S., Blinov,M.L., Redondo,A. and Wofsy,C. (2002) Modeling the early signaling events mediated by FcεRI. *Mol. Immunol.* **38**, 1213-1219.

Goldstein,B., Faeder,J.R. and Hlavacek,W.S. (2004) Mathematical and computational models of immune-receptor signalling. *Nat. Rev. Immunol.* **4**, 445-456.

Hlavacek,W.S., Faeder,J.R., Blinov,M.L., Perelson,A.S. and Goldstein,B. (2003) The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* **84**, 783-794.

Hucka,M., Finney,A., Sauro,H.M., Bolouri,H., Doyle,J.C., Kitano,H. and the rest of the SBML forum (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524-531.

Le Novère,N. and Shimizu,T.S. (2001) STOCHSIM: modelling of stochastic biomolecular processes. *Bioinformatics* **17**, 575-576.

Shapiro,B.E., Levchenko,A., Meyerowitz,E.M., Wold,B.J. and Mjolsness,E.D. (2003) Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* **19**, 677-678.

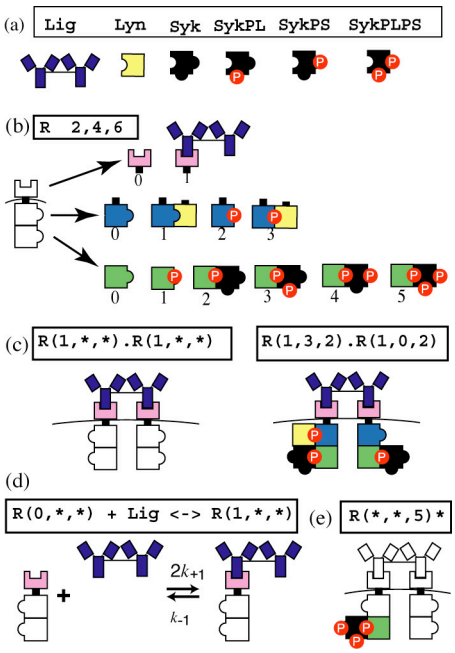

Fig. 1. Illustrated declarations in the input file (`fceri_net.in`) that specifies the model and output functions of Faeder et al. (2003). Boxes enclose text of the input file. (a) Declarations of six individual molecular species. (b) A multi-state species declaration of 48 individual molecular species that contain one receptor (R). Each of these species is characterized by three domains, which have two, four, and six possible states. (c) Declaration of complexes that contain two receptors (left) and a reference to one of the 300 individual molecular species in this class (right). (d) The reaction rule for ligand-receptor binding, which implies 24 distinct forward reactions and the same number of reverse reactions. All forward (reverse) reactions are assigned the rate constant k_{+1} (k_{-1}). (e) Declaration of an output function, a weighted sum of 98 concentrations, used to calculate the total concentration of autophosphorylated Syk.